



# **Canadian's Views about Using Big Data in Health Research from a National Online Survey: A Partnership of Consumer- Patients and Researchers**

Journal:	<i>CMAJ Open</i>
Manuscript ID	CMAJOpen-2018-0105
Manuscript Type:	Survey
Date Submitted by the Author:	11-Jul-2018
Complete List of Authors:	McCormick, Natalie; The University of British Columbia, Faculty of Pharmaceutical Sciences; Arthritis Research Canada, Hamilton, Clayon; The University of British Columbia, Department of Physical Therapy; Arthritis Research Canada, Koehn, Cheryl; Arthritis Consumer Experts English, Kelly; Arthritis Research Canada, Stordy, Allan; Canadian Skin Patient Alliance Li, Linda; University of British Columbia, Physical Therapy; Arthritis Research Canada,
Keywords:	Health services research, Knowledge translation
More Detailed Keywords:	Big data
Abstract:	<p>Background: Health research using "big data" (large sets of routinely-collected healthcare information) benefits patients and society, but growing public concerns about personal information being accessed for unintended purposes could erode trust and impose barriers to this work. We sought Canadians' views on big data in health research.</p> <p>Methods: Researchers and consumer-patient leaders of three joint and skin disease organisations partnered to develop and distribute a bilingual, online survey open to all Canadian adults. Survey asked respondents' initial perceptions about big data in health research, then (after providing some background information) asked their views on specific topics (i.e. benefits of big data, data access/privacy) and ongoing perceptions and educational needs.</p> <p>Results: 151 completed the survey (117=77% female; 47% aged 50-69 years). 101 (67%) had a chronic disease. At the start, 79% felt positively about use of big data for health research. Respondents ranked the ability to study large numbers of people (selected by 73%) and long-term treatment effects and rare events (76%) as the top benefits of using big data. De-identifying personal information was the most important privacy measure (selected by 89%) and 67% wanted to learn more about data</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

	<p>stewards granting access to data. At the end of the survey (after viewing background information about big data), 93% felt positively about big data (vs. 79% at the start) but only 58% were confident about privacy and security measures in place.</p> <p>Interpretation: More education, especially about access and privacy controls, may enhance public trust about using big data in health research.</p>

SCHOLARONE™  
Manuscripts

Confidential

**Canadians' Views about Using Big Data in Health Research from a National Online Survey: A Partnership of Consumer-Patients and Researchers**

Natalie McCormick, PhD<sup>1</sup>

Clayton B. Hamilton, PhD<sup>1, 2</sup>

Cheryl L. Koehn, BA<sup>3</sup>

Kelly English<sup>4</sup>

Allan Stordy<sup>5</sup>

Linda C. Li, BSc(PT), PhD<sup>1, 2</sup>

<sup>1</sup>Arthritis Research Canada, 5591 No. 3 Road, Richmond BC Canada V6X 2C7

<sup>2</sup>Department of Physical Therapy, The University of British Columbia, Vancouver BC Canada

<sup>3</sup>Arthritis Consumer Experts, #210 – 1529 West 6th Avenue, Vancouver BC Canada V6J 1R1

<sup>4</sup>Arthritis Patients Advisory Board, Arthritis Research Canada, 5591 No. 3 Road, Richmond BC Canada V6X 2C7

<sup>5</sup>Canadian Skin Patient Alliance, Calgary AB Canada

**Article Type:** Survey

**Word Count (excluding title page, abstract, references, tables and figures):** 2,496

**Correspondence to:** Dr. Linda C. Li: [lli@arthritisresearch.ca](mailto:lli@arthritisresearch.ca)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**ABSTRACT**

**Background:**

Health research using “big data” (large sets of routinely-collected healthcare information) benefits patients and society, but growing public concerns about personal information being accessed for unintended purposes could erode trust and impose barriers to this work. We sought Canadians’ views on big data in health research.

**Methods:**

Researchers and consumer-patient leaders of three joint and skin disease organisations partnered to develop and distribute a bilingual, online survey open to all Canadian adults. Survey asked respondents’ initial perceptions about big data in health research, then (after providing some background information) asked their views on specific topics (i.e. benefits of big data, data access/privacy) and ongoing perceptions and educational needs.

**Results:**

151 completed the survey (117=77% female; 47% aged 50-69 years). 101 (67%) had a chronic disease. At the start, 79% felt positively about use of big data for health research. Respondents ranked the ability to study large numbers of people (selected by 73%) and long-term treatment effects and rare events (76%) as the top benefits of using big data. De-identifying personal information was the most important privacy measure (selected by 89%) and 67% wanted to learn more about data stewards granting access to data. At the end of the survey (after viewing background information about big data), 93% felt positively about big data (vs. 79% at the start) but only 58% were confident about privacy and security measures in place.

**Interpretation:**

More education, especially about access and privacy controls, may enhance public trust about using big data in health research.

## INTRODUCTION

“Big data” is a mainstream media term to describe large repositories of information. In the context of Canadian health research, information routinely collected on behalf of provincial health ministries and other public bodies have advanced our knowledge of the burden(1) and risks(2) of diseases, long-term harms and benefits of treatments(3), and drivers of healthcare costs(4). As noted in a 2015 *CMAJ* editorial(5), a crucial advantage of these data is they cover Canadians of all ages, ethnicities, and sociodemographic groups, which makes findings more applicable to the entire Canadian population, including groups who tend not to participate in studies.

For those closely involved in Canadian health research with big data, these are exciting times. Researchers have access to a growing array of publicly-collected datasets covering healthcare utilisation, workplace safety, immigration, and early childhood development, as well as electronic medical records (EMR) data(6) and cancer and perinatal registries. Analyses of these data can improve our understanding of long-term health outcomes and the influence of social and behavioural factors. This research can, in turn, guide the treatment decisions made by individual patients and the development of interventions to address disparities at the population-level. For example, by linking 17 sources of data, including vital statistics and administrative data, laboratory databases, disease registries, and EMRs(7), Ontario researchers were able to identify major contributors to regional variation in cardiovascular event rates(8) and determine that high-intensity statin therapy offered minimal added-benefits to elderly patients(9).

However, the majority of Canadians may not be familiar with these advantages. Moreover, recent security breaches at major companies(10–12) and the Facebook-Cambridge Analytica data scandal(13) have left many worried about their personal information being accessed for unintended purposes. While this is a valid concern, a more balanced discussion is needed, particularly given the potential societal benefits of using big data. This would ensure that negative reporting in the public media, and misconceptions about what information can be accessed by Canadian health researchers, do not dominate the public discourse and place undue barriers on research of this type.

Among those leading this discussion are patients themselves, and there is growing recognition in Canada, including by the *CMAJ* group(14), of the value of patient-oriented research that focuses on patient-identified

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

priorities and patient engagement on research teams(15). Leaders of national consumer-patient organisations, including Arthritis Consumer Experts, the Arthritis Patient Advisory Board at Arthritis Research Canada, and the Canadian Skin Patient Alliance, are co-investigators on a multidisciplinary Canadian Institutes of Health Research (CIHR)-funded team grant(16) aiming to advance knowledge about prevention, burden, and management of serious complications in chronic inflammatory diseases. As part of their mandate to engage the public about their views on health research topics, the consumer-patient investigators identified a need to ascertain and enhance the public’s understanding about using large, ‘real-world’ datasets for research, and this was supported by the researchers on the team. Little is known about the general Canadian population’s views on this topic, with prior studies focussing on a single issue (namely, consent preferences for the research use of medical records data(17–19)) or the views of patients with specific diagnoses(20–22). Therefore, the researchers and consumer-patient investigators co-developed and distributed an online survey with the aim of ascertaining the general population’s views about the use of big data in Canadian health research.

Confidential

## METHODS

### Setting

Data were collected via a self-completed online survey open to individuals aged  $\geq 18$  years across Canada. The survey, titled “Is ‘Big Data’ a big deal for health care in Canada?”, was open for completion in English and French from January 17 through August 15, 2017.

### Design

Recruitment was carried out online, mainly through the websites, e-mail lists, and social media channels of the consumer-patients’ affiliate organisations. These efforts were supplemented by direct e-mails, social media (e.g., Twitter, Facebook), and other word-of-mouth communication between the grant investigators and their colleagues at health research institutes and patient advocacy organisations throughout Canada. The recruitment notices contained a link to the web-based survey hosted on The University of British Columbia’s FluidSurveys platform(23); all collected data remained in Canada. Before starting the survey, individuals were required to review a consent page and give their consent to participate. The consent page named the Principal Investigator and described the title and purpose of the study, and how and where survey responses would be stored. All questions were optional, no personal identifiers were collected, and no incentives were offered.

### Sources of data/measures

The initial version of the survey was co-developed by the lead consumer-patient investigator (CK) and two research trainees (NM and CH) experienced in health services research and survey design. A group of informed consumers, including the three consumer-patient co-authors (CK, KE, and AS), provided iterative feedback on the survey content with a primary goal of making it both understandable and interesting to the general public.

The survey was formulated in English, tested to ensure it could be completed within the allotted time (15 minutes), then translated into French by a professional translator. A copy is available in the Appendix. Given the many types of big data available for use in health research, and heterogeneity in their purpose, coverage, and privacy features, addressing all types in a single survey would be too burdensome. Thus, the survey focussed on the administrative datasets collected by public bodies (i.e. provincial/federal Ministries and agencies)(24) and covering

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

nearly all legal residents of each province and territory. Access to these publicly-collected datasets is controlled by data stewards(24), officials designated by the public bodies to adjudicate researchers’ requests for de-identified subsets of these data.

The 25-question survey consisted of six core sections along with a preamble which introduced respondents to big data and purpose of the survey. The first section asked about their familiarity and initial perceptions about the use of big data in health research. Over four subsequent sections, respondents were queried about their views on specific topics, including advantages of using big data for health research (overall, and in Canada specifically), types of health research questions that can be answered, and data access and privacy measures. In the final section, respondents were queried further about their perceptions of big data and interest in learning more following completion of the survey. The survey was primarily designed to ascertain respondents’ views, but we provided a small amount of educational information alongside the questions since we expected most respondents would be unfamiliar with big data. For example, when asking which privacy measures they felt were most important, we provided a description of each. We posed the same question at the start and end of the survey (“In general, how do you feel about the use of Big Data for health research?”) to explore whether respondents’ perceptions changed as they completed it.

**Statistical analysis**

Only submitted responses (where the respondent formally submitted the survey at the end of the last page) were analysed; missing responses to individual questions within these submissions were permitted. For each question, we calculated the percentage of respondents selecting each item. Since respondents were often asked to select multiple items from a list, the sum of percentage-frequencies could exceed 100%. Analysis were generated using the SAS software package, version 9.5 (Cary, North Carolina, USA).

**Ethics approval**

The study received ethics approval from the Behavioural Research Ethics Board at The University of British Columbia (#H16-02745).



## RESULTS

Two-hundred-and-thirty individuals provided consent, of whom 151 (66%) submitted responses and were included in the analysis. Of the 79 responses that were not submitted, the majority (n=47) were from individuals who consented to participate but never started the survey. More than three-quarters of respondents (n=117/151=78%) were female, nearly half (47%) were aged 50-69 years, and 28% were aged 30-49 years (Table 1). Most lived in British Columbia (BC) (56%) or Ontario (26%), and 58% had completed a university degree. Two-thirds (n=101) indicated they had a chronic disease.

### Initial knowledge and perceptions

Responses to section one showed more than three-quarters (79%) felt positively about the use of big data for health research, while 20% did not know, one respondent felt negatively, and one declined to answer. Approximately 95% had heard of the term “electronic health/medical record”, but only 58% knew the term “administrative health database” or “administrative data”.

### Perceived uses and advantages

Table 2 shows the results for sections two and three. Respondents selected the ability to study large numbers of people (selected by 73%) and long-term effects and rare events (selected by 75%) as the most important benefits of using big data. The benefits they most-wanted to learn about were studying long-term effects/rare events and potentially-harmful treatments (each selected by 64%). Similarly, the most important research questions to answer with big data involved the long-term harms and benefits of a particular treatment and complications of a particular disease (selected by 52-55%).

### Access and privacy

As shown in Table 3, the need to apply for the use of research data (selected by 62% of respondents) and obtain approval from university research ethics boards (selected by 59%) were the top-ranked data access controls, while two-thirds wanted to learn more about the role of the data stewards in granting access. De-identifying personal information from big data sets was selected by 89% of respondents as one of the most important privacy

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

measures, followed by mandating researchers to complete privacy training and sign confidentiality agreements (selected by 58%).

**Ongoing perceptions and next steps**

When questioned further about their perceptions of big data and educational interests, approximately 91% of respondents thought the provinces should promote big data for health research, and 93% were very or somewhat willing to have their de-identified data used by Canadian health researchers. Furthermore, when asked the same question on their overall feelings about big data, more felt positively at this point in the survey (141/151=93%) than at the start (119/151=79%). Even still, only 58% were confident about the privacy and security measures in place.

The top concern was insurance companies accessing data (selected by 59%) while the potential costs of collecting and overseeing the data were not a major concern (selected by 10%). Respondents' preferred mode for learning more about big data was via websites (selected by 88%), with little interest in receiving materials through the mail.

## INTERPRETATION

We aimed to ascertain the Canadian general population's views on the use of routinely-collected datasets in health research. Upon completing the online survey, more than 90% of respondents felt positively about big data, with the most appealing features being the ability to better understand the long-term benefits and harms of treatments. As well, consistent with reports from other countries(25–30), respondents placed high importance on the data being de-identified. However, while 93% were at least somewhat willing to have their de-identified information used by Canadian health researchers, far fewer were confident about the privacy and security procedures.

One of our most compelling and unexpected findings was how respondents' views changed during the survey: 79% felt positively about big data at the beginning, and 93% felt positively at the end. Response bias may have contributed, if respondents thought they *should* feel more positively by the end. However, the change may also have been due to the small amount of educational information provided alongside the questions. Although this hypothesis should be tested in future surveys for the public, support is provided by studies conducted in the United Kingdom(31,32) and New Zealand(33) where participants did report being more comfortable about the use of health data for research after receiving more information.

This potentially influential role of education is important given our respondents' desire to learn more about specific topics and their concerns about data access and privacy. Though generally supportive of big data in health research, less than 60% were confident about the privacy and security procedures in place. Ideally, a much larger percentage would be feeling confident about the privacy and security procedures for publicly-collected data in Canada. This suggests more work is needed to increase public awareness about big data and diminish lingering concerns. It is promising to think this could be accomplished through the provision of educational information like that provided within the survey. Some lay summaries, videos, and other educational materials are already available through organisations like Population Data BC(34) and Ontario's Institute for Clinical Evaluative Sciences(35) that facilitate research with big data. The public is likely not familiar with these organisations, but they could partner with consumer-patient groups to raise awareness about their websites and resources and assist in disseminating understandable, lay-language findings from their projects.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Existing resources could be supplemented by new ones tailored to what our respondents valued most about big data in health research and wanted to learn more about. For example, showcasing examples of Canadian studies where administrative data were used to assess complications(36–40) and the long-term effects of medications(41–45) in patients with chronic diseases would help convey the benefits of big data in a meaningful way. Similarly, providing more information about the data stewards’ role in adjudicating data access requests and imposing other conditions (i.e. privacy training, review of research outputs before publication) may help the public feel more confident about the privacy measures in place.

Our study was the product of a dynamic partnership between consumer-patients and researchers who worked together to develop the survey, recruit participants, and interpret findings. This partnership should continue into the dissemination stages, with researchers and patients co-developing the above-mentioned educational materials and sharing their own perspectives about big data with the public. Prior work(32,33,46) suggests members of the public benefit from hearing the perspectives of those involved and impacted by this type of health research, including ethicists, informaticians, researchers, clinicians, public health leaders, and patients.

**Limitations**

We realise findings from our online survey may not be generalisable to the entire Canadian population. Half of respondents lived in BC, and, as is typical with online surveys, our sample had a relatively-high level of education (58% university graduates). Moreover, while the survey was open to all Canadian adults, recruitment was carried out mainly through patient research and advocacy groups and our convenience sample (67% of whom had a chronic illness) may be more interested and willing to participate in health research than the public at-large. Finally, this survey focussed on features of administrative data and Canadians’ views on other sources of big data such as EMRs, biospecimens, and mobile health apps should be ascertained in future studies.

**Conclusion**

The vast majority of respondents who completed this first-of-its-kind Canadian survey felt positively about using big data in health research, and this increased after viewing information within the survey itself. However, many lacked confidence in the access and privacy controls for these publicly-collected datasets. As new sources of

1  
2  
3 health information become available for linkage and analysis, public deliberation programs like that now underway  
4  
5 in BC(47) may help ensure the country's data access and governance policies are in alignment with Canadians'  
6  
7 views. In the meantime, educational resources incorporating the publics' concerns and learning interests are needed  
8  
9 to enhance awareness and trust in health research using big data, and the resultant benefits for population health and  
10  
11 patient care.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**ACKNOWLEDGEMENTS**

We thank all of the individuals who participated in the survey. We also thank Lynn Nowoselski, M.P.H, who coordinated this project.

This study was funded by “PRECISION: Preventing Complications from Inflammatory Skin, Joint and Bowel Conditions” a Team Grant from the Canadian Institutes of Health Research, Canada (THC-316595). Natalie McCormick was funded by a Doctoral Research Award from the Canadian Institutes of Health Research. Clayon Hamilton is funded by a Michael Smith Foundation for Health Research (MSFHR) Trainee Award in Health Services Research and a Canadian Institutes of Health Services Research Fellowship. Linda Li is supported by the Canada Research Chair Program, the Harold Robinson/Arthritis Society Chair in Arthritic Diseases, and the Michael Smith Foundation for Health Research (MSFHR) Scholar Award.

**CONTRIBUTOR STATEMENT**

Clayon Hamilton, Cheryl Koehn, Kelly English, Allan Stordy, and Linda Li conceived the study. Clayon Hamilton analysed the data. Natalie McCormick and Clayon Hamilton drafted the manuscript.

All of the authors contributed substantially to the study design and interpretation of data, revised the manuscript critically for important intellectual content, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

## REFERENCES

1. Benchimol EI, Bernstein CN, Bitton A, Carroll MW, Singh H, Otley AR, et al. Trends in Epidemiology of Pediatric Inflammatory Bowel Disease in Canada: Distributed Network Analysis of Multiple Population-Based Provincial Health Administrative Databases. *Am J Gastroenterol*. 2017 Jul;112(7):1120–34.
2. Lega IC, Wilton AS, Austin PC, Fischer HD, Johnson JA, Lipscombe LL. The temporal relationship between diabetes and cancer: A population-based study: Population-Based Study of Diabetes and Cancer. *Cancer*. 2016 Sep 1;122(17):2731–8.
3. Rawson NSB, Downey W, Maxwell CJ, West R. 25 years of pharmacoepidemiologic innovation: the Saskatchewan health administrative databases. *J Popul Ther Clin Pharmacol J Ther Popul Pharmacologie Clin*. 2011;18(2):e245-249.
4. Fitzpatrick T, Rosella LC, Calzavara A, Petch J, Pinto AD, Manson H, et al. Looking Beyond Income and Education: Socioeconomic Status Gradients Among Future High-Cost Users of Health Care. *Am J Prev Med*. 2015 Aug;49(2):161–71.
5. Lougheed T. Rethinking barriers to big data. *Can Med Assoc J*. 2015 Jan 6;187(1):E1–E1.
6. Tu K, Mitiku TF, Ivers NM, Guo H, Lu H, Jaakkimainen L, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *Am J Manag Care*. 2014;20(1):e15-21.
7. Tu JV, Chu A, Donovan LR, Ko DT, Booth GL, Tu K, et al. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes*. 2015 Mar;8(2):204–12.
8. Tu JV, Chu A, MacLagan L, Austin PC, Johnston S, Ko DT, et al. Regional variations in ambulatory care and incidence of cardiovascular events. *Can Med Assoc J*. 2017 Apr 3;189(13):E494–501.
9. Alter DA, Tu JV, Koh M, Jackevicius CA, Austin PC, Rezai MR, et al. Projected Real-World Effectiveness of Using Aggressive Low-Density Lipoprotein Cholesterol Targets Among Elderly Statin Users Following Acute Coronary Syndromes in Canada. *J Am Heart Assoc*. 2018 May;7(10):e007535.
10. McCooley P. Quebecer charged in PC Plus breach, collectors told to change password | Ottawa Citizen. *Ottawa Citizen* [Internet]. 2017 Feb 22 [cited 2018 Apr 17]; Available from: <http://ottawacitizen.com/news/local-news/laval-man-charged-in-pc-plus-breach-collectors-urged-to-fortify-password>
11. Equifax says more than 19,000 Canadians affected by security breach | CBC News [Internet]. 2017 [cited 2018 Apr 17]. Available from: <http://www.cbc.ca/news/business/equifax-canadians-affected-update-1.4424066>
12. Bell Canada alerts customers impacted by new data breach - National | Globalnews.ca [Internet]. 2018 [cited 2018 Apr 17]. Available from: <https://globalnews.ca/news/3982235/bell-canada-data-breach/>
13. Braga M. Facebook says more than 600,000 Canadians may have had data shared with Cambridge Analytica | CBC News [Internet]. 2018 [cited 2018 Apr 22]. Available from: <http://www.cbc.ca/news/technology/facebook-cambridge-analytica-600-thousand-canadians-1.4605097>
14. Patrick K, Kebbe M, Aubin D. A home for patient-oriented research. *Can Med Assoc J*. 2018 May 22;190(20):E607–E607.
15. Government of Canada. Strategy for Patient-Oriented Research - CIHR [Internet]. 2010 [cited 2018 May 16]. Available from: <http://www.cihr-irsc.gc.ca/e/41204.html>

16. PRECISION | Arthritis Research Canada [Internet]. [cited 2018 Apr 21]. Available from: <http://www.arthritisresearch.ca/precision>

17. Nair K, Willison D, Holbrook A, Keshavjee K. Patients' consent preferences regarding the use of their health information for research purposes: a qualitative study. *J Health Serv Res Policy*. 2004 Jan;9(1):22–7.

18. Willison DJ, Swinton M, Schwartz L, Abelson J, Charles C, Northrup D, et al. Alternatives to project-specific consent for access to personal information for health research: Insights from a public dialogue. *BMC Med Ethics* [Internet]. 2008 Dec [cited 2018 Apr 17];9(1). Available from: <http://bmcmethics.biomedcentral.com/articles/10.1186/1472-6939-9-18>

19. Willison DJ, Schwartz L, Abelson J, Charles C, Swinton M, Northrup D, et al. Alternatives to Project-specific Consent for Access to Personal Information for Health Research: What Is the Opinion of the Canadian Public? *J Am Med Inform Assoc*. 2007 Nov 1;14(6):706–12.

20. Page SA, Manhas KP, Muruve DA. A survey of patient perspectives on the research use of health information and biospecimens. *BMC Med Ethics* [Internet]. 2016 Dec [cited 2018 May 17];17(1). Available from: <http://bmcmethics.biomedcentral.com/articles/10.1186/s12910-016-0130-4>

21. Terry AL, Chesworth BM, Bourne RB, Stolee P, Speechley M. Joint replacement recipients' views about health information privacy. *Health Expect*. 2015 Oct;18(5):1519–29.

22. Page SA, Mitchell I. Patients' opinions on privacy, consent and the disclosure of health information for medical research. *Chronic Dis Can*. 2006;27(2):60–7.

23. Survey Tool (FluidSurveys) | UBC Information Technology [Internet]. [cited 2018 Jun 21]. Available from: <https://it.ubc.ca/services/teaching-learning-tools/survey-tool-fluidsurveys>

24. Glossary | www.popdata.bc.ca [Internet]. [cited 2018 Jun 12]. Available from: <https://www.popdata.bc.ca/BCDataDeliberation/glossary>

25. Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics*. 2015 Nov 17;16(1):79.

26. Buckley BS, Murphy AW, MacFarlane AE. Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public. *J Med Ethics*. 2011 Jan 1;37(1):50–5.

27. Luchenski SA, Reed JE, Marston C, Papoutsis C, Majeed A, Bell D. Patient and public views on electronic health records and their uses in the United kingdom: cross-sectional survey. *J Med Internet Res*. 2013 Aug 23;15(8):e160.

28. Riordan F, Papoutsis C, Reed JE, Marston C, Bell D, Majeed A. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *Int J Med Inf*. 2015 Apr;84(4):237–47.

29. Zalin A, Papoutsis C, Shotliff K, Majeed A, Marston C, Reed J. The use of information for diabetes research and care: patient views in West London. *Pract Diabetes*. 2016 Apr;33(3):81-86a.

30. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. *Int J Med Inf*. 2006 Jul;75(7):530–41.

31. Hill EM, Turner EL, Martin RM, Donovan JL. “Let's get the best quality research we can”: public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study.



- BMC Med Res Methodol [Internet]. 2013 Dec [cited 2018 May 14];13(1). Available from: <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-72>
32. Tully MP, Bozentko K, Clement S, Hunn A, Hassan L, Norris R, et al. Investigating the Extent to Which Patients Should Control Access to Patient Records for Research: A Deliberative Process Using Citizens' Juries. *J Med Internet Res*. 2018 Mar 28;20(3):e112.
  33. Parkin L, Paul C. Public good, personal privacy: a citizens' deliberation about using medical information for pharmacoepidemiological research. *J Epidemiol Community Health*. 2011 Feb 1;65(2):150–6.
  34. Population Data BC | [www.popdata.bc.ca](http://www.popdata.bc.ca) [Internet]. [cited 2018 Apr 17]. Available from: <https://www.popdata.bc.ca/>
  35. ICES | Institute for Clinical Evaluative Sciences [Internet]. [cited 2018 Apr 17]. Available from: <https://www.ices.on.ca/>
  36. Feldman DE, Vinet É, Bérard A, Duffy C, Hazel B, Meshefedjian G, et al. Heart Disease, Hypertension, Gestational Diabetes Mellitus, and Preeclampsia/Eclampsia in Mothers With Juvenile Arthritis: A Nested Case-Control Study: Maternal Outcomes in Juvenile Arthritis. *Arthritis Care Res*. 2017 Feb;69(2):306–9.
  37. Marrie RA, Patten SB, Greenfield J, Svenson LW, Jette N, Tremlett H, et al. Physical comorbidities increase the risk of psychiatric comorbidity in multiple sclerosis. *Brain Behav*. 2016 Sep;6(9):e00493.
  38. Kuenzig ME, Barnabe C, Seow CH, Eksteen B, Negron ME, Rezaie A, et al. Asthma Is Associated With Subsequent Development of Inflammatory Bowel Disease: A Population-based Case–Control Study. *Clin Gastroenterol Hepatol*. 2017 Sep;15(9):1405–1412.e3.
  39. Levesque A, Lachaine J, Bissonnette R. Risk of myocardial infarction in Canadian patients with psoriasis: a retrospective cohort study. *J Cutan Med Surg*. 2013 Dec;17(6):398–403.
  40. Solomon DH, Love TJ, Canning C, Schneeweiss S. Risk of diabetes among patients with rheumatoid arthritis, psoriatic arthritis and psoriasis. *Ann Rheum Dis*. 2010 Dec;69(12):2114–7.
  41. Raymakers AJN, Sadatsafavi M, Sin DD, De Vera MA, Lynd LD. The Impact of Statin Drug Use on All-Cause Mortality in Patients With COPD. *Chest*. 2017 Sep;152(3):486–93.
  42. CAN-AIM, Moura CS, Abrahamowicz M, Beauchamp M-E, Lacaille D, Wang Y, et al. Early medication use in new-onset rheumatoid arthritis may delay joint replacement: results of a large population-based study. *Arthritis Res Ther* [Internet]. 2015 Dec [cited 2018 Apr 21];17(1). Available from: <http://arthritis-research.com/content/17/1/197>
  43. Dixon WG, Abrahamowicz M, Beauchamp M-E, Ray DW, Bernatsky S, Suissa S, et al. Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: a nested case–control analysis. *Ann Rheum Dis*. 2012 Jul;71(7):1128–33.
  44. Amiche MA, Lévesque LE, Gomes T, Adachi JD, Cadarette SM. Effectiveness of Oral Bisphosphonates in Reducing Fracture Risk Among Oral Glucocorticoid Users: Three Matched Cohort Analyses. *J Bone Miner Res*. 2018 Mar;33(3):419–29.
  45. Shah M, Avgil Tsadok M, Jackevicius CA, Essebag V, Eisenberg MJ, Rahme E, et al. Warfarin Use and the Risk for Stroke and Bleeding in Patients With Atrial Fibrillation Undergoing Dialysis. *Circulation*. 2014 Mar 18;129(11):1196–203.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

46. Douglas A, Ward HJT, Bhopal R, Kirkpatrick T, Sayed-Rafiq A, Gruer L. Is the linkage of census and health data justified? Views from a public panel of the Scottish Health and Ethnicity Linkage study. J Public Health. 2017 May 25;1–6.

47. BC Data Deliberation | [www.popdata.bc.ca](http://www.popdata.bc.ca) [Internet]. [cited 2018 Jun 14]. Available from: <https://www.popdata.bc.ca/BCDataDeliberation>

Confidential

Table 1: Respondents' Characteristics

	N (%)
<b>Total Respondents</b>	151
<b>Female Gender</b>	117 (77%)
<b>Age Group</b>	
18-29 years	21 (14%)
30-49 years	42 (28%)
50-69 years	71 (47%)
70-79 years	16 (11%)
≥ 80 years	1 (1%)
<b>Educational Attainment</b>	
High school or less	10 (7%)
Some community college, technical, trade, or vocational college	23 (15%)
Community college degree/diploma, or some university (but no degree)	30 (20%)
University degree or higher	87 (58%)
<b>Province/Territory of Residence<sup>a</sup></b>	
British Columbia	84 (56%)
Alberta	7 (5%)
Ontario	39 (26%)
Quebec	13 (9%)
Nova Scotia	2 (1%)
New Brunswick	3 (2%)
Prince Edward Island	1 (1%)
<b>Living with a Chronic Disease</b>	101 (67%)

Declined to answer: Gender (n=2), Education (n=1), Province/Territory (n=2), Chronic Disease (n=6)

<sup>a</sup>No responses were received from Saskatchewan, Manitoba, Newfoundland and Labrador, Northwest Territories, Yukon Territory, or Nunavut.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

Table 2: Responses on Reasons to Use Big Data for Health Research (Most- to Least-Frequently Selected)

Advantages of Using Big Data								
	1. Long-Term Effects and Rare Events	2. Large Numbers	3. Study Potentially-Harmful Treatments	4. General Population Comparisons	5. More Inclusive			
Most Important (select up to 3)	75.5%	72.8%	50.3%	46.4%	43.0%			
	1. Study Potentially-Harmful Treatments	2. Long-Term Effects and Rare Events	3. Large Numbers	4. General Population Comparisons	5. More Inclusive			
Want Additional Information About (select up to 3)	64.2%	63.6%	44.4%	41.7%	35.1%			
Advantages of using Big Data from Canada								
	1. Reflective of Canadian Health Care System	2. More Inclusive	3. Universal Prescription Medication Data	4. Reflective of Canadian Population				
Most Important (select up to 2)	66.2	63.6	35.8	27.8				
Topics to Study Using Big Data								
	1. Treatment Benefits	2. Treatment Harms	3. Disease Complications	4. Changes in Policy or Practice	5. Quality of Care	6. Cost-Effectiveness	7. Risk Factors for Disease	8. Disease Incidence and Prevalence
Most Important (select up to 3)	55.6%	55.0%	52.3%	43.7%	30.5%	27.8%	23.2%	9.9%

Expressed as the percentage selecting each response option; as multiple responses could be selected, the sum of percentage-frequencies exceeds 100%

Table 3: Responses on Data Access and Privacy and Security Controls (Most- to Least-Frequently Selected)

Data Access Controls					
	1. Must Apply for Data Access	2. Approval from Research Ethics Board	3. Approval from Data Stewards	4. Access Data for Limited Time	
<b>Most Important</b> (select up to 2)	62.3%	58.9%	51.0%	20.5%	
	1. Approval from Data Stewards	2. Approval from Research Ethics Board	3. Access Data for Limited Time	4. Must Apply for Data Access	
<b>Want Additional Information About</b> (select up to 2)	66.9%	46.4%	29.8%	25.2%	
Privacy and Security Controls					
	1. Data are De-Identified	2. Privacy Training and Confidentiality Agreement	3. Review of Research Outputs	4. Funding Agencies Cannot Access Data	5. No Access Outside Canada
<b>Most Important</b> (select up to 3)	89.4%	57.6%	43.7%	35.8%	35.1%

Expressed as the percentage selecting each response option; as multiple responses could be selected, the sum of percentage-frequencies exceeds 100%

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# Is ‘Big Data’ a big deal for health care in Canada?

---

## Is ‘Big Data’ a big deal for health care in Canada?

The term “Big Data” is used a lot these days, especially in health research. Generally speaking, “Big Data” are large and complex sets of data with information routinely collected across health care providers about the health and health care use of people across Canada. Researchers can use this data to learn about key health issues, such as the number of people in Canada living with a certain disease, complications experienced by people with those diseases, long-term harms and benefits of different treatments, and health care costs. There is growing interest in using Big Data in health research because it allows researchers to analyze more information than is available from patients attending a single hospital or clinic. What we can learn from Big Data could improve health policymaking in Canada, and could help patients and providers in making informed health care decisions. However, not much is known about patients’ and the public’s views about the use of Big Data for health research. For this reason, patient organizations and researchers in the PRECISION Network launched this survey to understand patients’ knowledge and opinions about the use of Big Data in Canadian health research. We want to hear your views, and sincerely appreciate the time you will give to answer each question to the best of your ability. The survey will take approximately 15 minutes to complete. The information you provide in this survey will remain strictly confidential and accessed only by members of the PRECISION study team. Your participation will be anonymous and your data will be pooled for analysis to ensure your complete privacy.

## About PRECISION

*PRECISION: Preventing complications from inflammatory skin, joint and bowel conditions* research network is studying a number of types of arthritis as well as skin disease, psoriasis, and two types of bowel disease: Crohn’s disease and ulcerative colitis. Why? Researchers believe these diseases share some very common features and complications and if health care providers got better at controlling the acute problems of these diseases, we would be left with chronic, low-grade, ongoing inflammation. The complications of these diseases now outweigh the original problems that have become treatable with better therapies. The PRECISION Network is researching what links complications (like heart attacks, stroke and others) to the inflammation, and ways to intervene to eliminate and prevent them.

## SECTION 1: WHAT IS BIG DATA?

### 1. Before starting this survey, were you familiar with the use of Big Data for health research?

- ☐ Yes
- ☐ No
- ☐ Decline to answer

### 2. In general, how do you feel about the use of Big Data for health research?

- ☐ Positively
- ☐ Negatively
- ☐ I Don't Know
- ☐ Decline to answer

### 3. Who do you think uses the health research findings from Big Data? For each item, please select Yes, No, or Don't Know:

	Yes	No	Don't Know
Physicians and other health care providers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Patients	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
University researchers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consumer-patient advocacy or research groups	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Groups that make guidelines and recommendations for clinical practice (for example, Canadian Medical Association)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Health policy-makers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provincial or federal governments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Big Data for health research can include electronic health records, administrative health databases, and medical laboratory results.

4. Have you heard of the term “electronic health record” or “electronic medical record”?

- ☐ Yes
- ☐ No
- ☐ Decline to answer

5. Have you heard of the term “administrative health database” or “administrative data”?

- ☐ Yes
- ☐ No
- ☐ Decline to answer

The terms “administrative health database” and “administrative data” refer to information collected daily about the delivery of healthcare to residents in a Canadian province or territory. Healthcare providers and staff, such as pharmacists and billing clerks, collect this information as part of their jobs.

SECTION 2: WHY BIG DATA?

Big Data can be more useful than other types of health data, because it brings together information collected from all residents of a region (e.g., city, health authority, province, or country) receiving care in a variety of healthcare settings.

6. Below are some benefits of using Big Data for health research. Please select up to three (3) that you feel are the most important:

- ☐ **Large Numbers:** Big Data makes it easy to study many people, including those with multiple health conditions.
- ☐ **Long-Term Effects and Rare Events:** Big Data uses information collected on lots of people, over many years, so long-term effects and rare events can be studied.
- ☐ **More Inclusive:** Big Data makes research findings relevant to more people, by using information from everyone who uses health care services.



- ☐ **General Population Comparisons:** Big Data allows for better quality source information since people who have a certain disease or receive a certain treatment are compared to others from the general population.
- ☐ **Studying Potentially-Harmful Treatments:** Big Data is a great way to study potentially harmful treatments because it accumulates information on what is actually happening.

## 7. Please choose up to three (3) benefits that you would like to know more about:

- ☐ **Large Numbers**
- ☐ **Long-Term Effects and Rare Events**
- ☐ **More Inclusive**
- ☐ **General Population Comparisons**
- ☐ **Studying Potentially-Harmful Treatments**

## SECTION 3: WHAT CAN BIG DATA DO FOR ME?

Big Data is used to study many important issues related to health policy and patient care.

## 8. Please choose up to three (3) issues that are the most important to you:

- ☐ **Benefits** of using a particular treatment
- ☐ **Harms** of using a particular treatment
- ☐ **Complications or other health problems** that affect people with certain diseases
- ☐ Factors that **affect your chances** of developing a disease
- ☐ **Quality of care** provided to people with a certain disease
- ☐ Number of people **diagnosed or living with** a certain disease
- ☐ **Cost-effectiveness** of new medications, programs, or treatment strategies
- ☐ How a **change in policy or patient care** affects patient health outcomes

## SECTION 4: WHY USE BIG DATA FROM CANADA?

Health researchers across Canada, and in many other countries, use administrative health databases and other types of Big Data.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**9. Below we list some of the many benefits of using Big Data from Canada for health research. Please select up to two (2) benefits that you feel are most important:**

- ☐ **Data on all Canadians:** Because Canada has a public health care system, its provincial and territorial healthcare databases contain information on all residents, regardless of age or employment status. This makes the data more inclusive, and less biased.
- ☐ **Universal prescription medication data:** In British Columbia and some other Canadian provinces, data are available on **prescription medications for all residents**, regardless of their age or drug coverage.
- ☐ **Reflective of the Canadian population:** Using data collected from a large Canadian population can provide information that is directly **applicable to those living here**.
- ☐ **Reflective of the Canadian health care system:** Analyzing **data on care delivered in the Canadian system** will provide the best information on how to improve care for Canadians.

**SECTION 5: TRUST, PRIVACY, AND SECURITY**

Administrative data have been used for health research in Canada for many years. There are a number of measures in place to ensure the responsible use of data.

**10. Please select up to two (2) measures that you feel are the most important:**

- ☐ Researchers **must apply** to access data for each project
- ☐ Researcher must get **university research ethics board approval** before accessing the data
- ☐ All applications must be approved by the **relevant Data Stewards**, who are the individuals responsible for the databases
- ☐ Researchers may only use the data for a **limited period of time**

**11. Below is the same list. Please choose up to two (2) that you would like to know more about:**

- ☐ Researchers **must apply** to access data for each project
- ☐ Researcher must get **university research ethics board approval** before accessing the data
- ☐ All applications must be approved by the **relevant Data Stewards**, who are the individuals responsible for the databases
- ☐ Researchers may only use the data for a **limited period of time**

Several measures are in place to enhance the security of the information and protect your privacy.

**12. Please select up to three (3) that you feel are most important:**

- ☐ Research datasets **do not include any identifying information** about patients or their family members – no names, addresses, phone numbers, Personal Health Numbers, or Social Insurance Numbers
- ☐ Researchers complete **Privacy Training and sign a confidentiality agreement** before accessing the data
- ☐ Funding agencies **do not have access** to the data
- ☐ Data **cannot be accessed outside of Canada**
- ☐ **All publications, presentations, and other reports** about the research **are reviewed** to ensure all privacy regulations have been followed

**SECTION 6: NEXT STEPS**

**13. Now that you've gone through this survey, please tell how much you agree or disagree with each of the following reasons for using Big Data in health research:**

	Strongly Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Strongly Disagree
Allows information and experiences of <b>all Canadians</b> to be included	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be used to answer questions about <b>real-life exposures and behaviours</b> that can't be studied in a <b>clinical trial</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be used to assess the <b>long-term effects</b> of treatments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Represents the <b>experiences of Canadians</b> receiving care in <b>our own health care system</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Data are already collected</b> for	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

March 14 2017- Version 4

1  
2  
3  
4 billing and other administrative  
5 purposes, **using taxpayer**  
6 **dollars**, and **shouldn't go to**  
7 **waste**  
8  
9  
10

11 **14. In general, having gone through this survey, how do you feel about the use**  
12 **of Big Data for health research?**  
13

- 14 ☐ Positively  
15  
16 ☐ Negatively  
17  
18 ☐ I Don't Know  
19  
20 ☐ Decline to answer  
21  
22

23 **15. Do you think Canadian provinces should promote the use of Big Data**  
24 **(without names or other identifying details) for health research?**  
25

- 26 ☐ Yes  
27  
28 ☐ No  
29  
30 ☐ I Don't Know  
31  
32 ☐ Decline to answer  
33  
34

35 **16. Do you feel confident about the privacy and security procedures in place for**  
36 **Big Data?**  
37

- 38 ☐ Yes  
39  
40 ☐ No  
41  
42 ☐ I Don't Know  
43  
44 ☐ Decline to answer  
45  
46

47 **17. How willing are you to have your data (without names or other identifying**  
48 **details) used by Canadian health researchers?**  
49

- 50  
51 ☐ Very willing  
52  
53 ☐ Somewhat willing  
54  
55 ☐ Neutral  
56  
57 ☐ Not too willing  
58

59 March 14 2017- Version 4  
60

- ☐ Not at all willing  
☐ Decline to answer

**18. Please select up to two (2) of the most important concerns you still have about Big Data use for health research:**

- ☐ Insurance companies accessing the data  
☐ Employers (or potential employers) accessing the data  
☐ Members of the public accessing the data  
☐ Access or storage of data outside Canada  
☐ Identity theft  
☐ Costs of collecting, storing, and overseeing the data  
☐ Other, please specify... \_\_\_\_\_  
☐ I have no concerns about Big Data  
☐ Decline to answer

**19. Through which modes would you like to learn more about Big Data in Canadian health research? Please select Yes or No for each:**

	Yes	No
Website	<input type="radio"/>	<input type="radio"/>
Online presentations or chat sessions	<input type="radio"/>	<input type="radio"/>
Direct e-mail	<input type="radio"/>	<input type="radio"/>
Printed materials by mail	<input type="radio"/>	<input type="radio"/>
Twitter	<input type="radio"/>	<input type="radio"/>
In-person presentations	<input type="radio"/>	<input type="radio"/>

1  
2  
3  
4 **Finally, please tell us a bit about yourself! Remember, all information you**  
5 **provide in this survey will remain confidential and will be accessed only by**  
6 **members of the PRECISION study team. No personal identifying information is**  
7 **being collected:**  
8  
9

10  
11  
12 **20. Are you:**

- 13 ☐ Male
- 14
- 15 ☐ Female
- 16
- 17 ☐ Other
- 18
- 19 ☐ Decline to answer
- 20

21  
22  
23 **21. In which Canadian province or territory do you live?**

- 24 ☐ Alberta
- 25
- 26 ☐ British Columbia
- 27
- 28 ☐ Manitoba
- 29
- 30 ☐ New Brunswick
- 31
- 32 ☐ Newfoundland and Labrador
- 33
- 34 ☐ Northwest Territories
- 35
- 36 ☐ Nova Scotia
- 37
- 38 ☐ Nunavut
- 39
- 40 ☐ Ontario
- 41
- 42 ☐ Prince Edward Island
- 43
- 44 ☐ Quebec
- 45
- 46 ☐ Saskatchewan
- 47
- 48 ☐ Yukon Territory
- 49
- 50 ☐ Decline to answer

51 **22. How old are you?**

- 52 ☐ 18-29 years
- 53
- 54 ☐ 30-49 years
- 55
- 56 ☐ 50-69 years

- ☐ 70-79 years
- ☐ 80 years or older
- ☐ Decline to answer

**23. What is the highest level of education you have completed?**

- ☐ Less than high school
- ☐ High school graduate or equivalent
- ☐ Some community college, technical, trade, or vocational college
- ☐ Community college degree/diploma, or some university (but no degree)
- ☐ University degree or higher
- ☐ Decline to answer

**24. We are interested in knowing whether people living with chronic health conditions have different views about Big Data than do other people. Have you been diagnosed by a health professional with a chronic health disease? By this, we mean a “long-term” disease that has already lasted, or is expected to last, 6 months or more.**

- ☐ Yes
- ☐ No
- ☐ Decline to answer

**25. Your input is important! In the box below, feel free to share any other thoughts or questions you have about the use of Big Data in Canadian health research:**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Thank you for taking part in the “Is ‘Big Data’ a big deal for healthcare in Canada?” survey, and helping the PRECISION Network consumer-patient organizations and researchers advance their work to benefit patients across Canada. If you would like to receive updates about this project, please provide your e-mail address here – your e-mail address will remain separate from your survey responses:**

Confidential



## Checklist for Reporting Results of Internet E-Surveys (CHRRIES)

Item Category	Checklist Item	Explanation	Page #
<b>Design</b>			
	Describe survey design	Describe target population, sample frame. Is the sample a convenience sample? (In "open" surveys this is most likely.)	Target population was Canadian adults: Page 5, paragraph 1.  Ours was a convenience sample: Page 10, paragraph 3, sentence 3.
<b>IRB approval and informed consent process</b>			
	IRB approval	Mention whether the study has been approved by an IRB.	IRB approval was granted: Bottom of page 6.
	Informed consent	Describe the informed consent process. Where were the participants told the length of time of the survey, which data were stored and where and for how long, who the investigator was, and the purpose of the study?	Before starting the survey, individuals were required to review a consent page and give their consent to participate. The consent page included information on the title and purpose of the study, Principal Investigator, and storage of data: Page 5, paragraph 2.  The expected length of time needed to complete the survey (15 minutes) was provided in the survey preamble: Page 5, paragraph 4, sentence 1.
	Data protection	If any personal information was collected or stored,	No personal

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Item Category	Checklist Item	Explanation	Page #
		describe what mechanisms were used to protect unauthorized access.	identifiers were collected: Page 5, end of paragraph 2.  All information collected from participants was stored in a password-protected, restricted-access folder in a network with bank security level data protection.
Development & pretesting			
	Development testing	State how the survey was developed, including whether the usability and technical functionality of the electronic questionnaire had been tested before fielding the questionnaire.	Page 5, "Sources of data/measures" subheading.
Recruitment process and description of the sample having access to the questionnaire			
	Open survey versus closed survey	An "open survey" is a survey open for each visitor of a site, while a closed survey is only open to a sample which the investigator knows (password-protected survey).	Open survey, all Canadian adults were eligible to participate: Page 5, paragraph 1, sentence 1.
	Contact mode	Indicate whether or not the initial contact with the potential participants was made on the Internet. (Investigators may also send out questionnaires by mail and allow for Web-based data entry.)	Recruitment was carried out online: Page 5, paragraph 2, sentence 1.
	Advertising the survey	How/where was the survey announced or advertised? Some examples are offline media (newspapers), or online (mailing lists – If yes, which ones?) or banner ads (Where were these banner ads posted and what did they look like?). It is important to know the wording of the announcement as it will heavily influence who chooses to participate. Ideally the survey announcement should be published as an appendix.	Advertised through the websites, e-mail lists, and social media channels of Canadian consumer-patient advocacy organisations and health research institutes: Page 5, paragraph 2, sentences 1 & 2.

Item Category	Checklist Item	Explanation	Page #
<b>Survey administration</b>			
	Web/E-mail	State the type of e-survey (eg, one posted on a Web site, or one sent out through e-mail). If it is an e-mail survey, were the responses entered manually into a database, or was there an automatic method for capturing responses?	Survey was posted on a website: Page 5, paragraph 2, sentence 3.
	Context	Describe the Web site (for mailing list/newsgroup) in which the survey was posted. What is the Web site about, who is visiting it, what are visitors normally looking for? Discuss to what degree the content of the Web site could pre-select the sample or influence the results. For example, a survey about vaccination on a anti-immunization Web site will have different results from a Web survey conducted on a government Web site	Recruitment is described on Page 5, paragraph 2, sentences 1 & 2.  Generalisability of the survey sample is discussed on Page 10, paragraph 3, sentences 1-3.
	Mandatory/voluntary	Was it a mandatory survey to be filled in by every visitor who wanted to enter the Web site, or was it a voluntary survey?	N/A – Voluntary survey advertised on multiple platforms; recruitment notices directed individuals to a separate website where the survey was posted.
	Incentives	Were any incentives offered (eg, monetary, prizes, or non-monetary incentives such as an offer to provide the survey results)?	No incentives were offered: Page 5, end of paragraph 2.
	Time/Date	In what timeframe were the data collected?	Data were collected from January 17 through August 15, 2017: Page 5, paragraph 1, sentence 2.
	Randomization of items or questionnaires	To prevent biases items can be randomized or alternated.	N/A – response items were not randomised.
	Adaptive questioning	Use adaptive questioning (certain items, or only conditionally displayed based on responses to other items) to reduce number and complexity of the questions.	N/A – there was no adaptive questioning.
	Number of Items	What was the number of questionnaire items per page? The number of items is an important factor for the	Depending on the length,

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Item Category	Checklist Item	Explanation	Page #
		completion rate.	approximately 1-3 questions per page. The survey was designed so each page could be viewed with minimal or no scrolling.
	Number of screens (pages)	Over how many pages was the questionnaire distributed? The number of items is an important factor for the completion rate.	Not available.
	Completeness check	It is technically possible to do consistency or completeness checks before the questionnaire is submitted. Was this done, and if “yes”, how (usually JavaScript)? An alternative is to check for completeness after the questionnaire has been submitted (and highlight mandatory items). If this has been done, it should be reported. All items should provide a non-response option such as “not applicable” or “rather not say”, and selection of one response option should be enforced.	N/A – no completeness checks were performed.  All questions were optional (page 5, end of paragraph 2) and “Decline to Answer” was included as a response option for most questions.
	Review step	State whether respondents were able to review and change their answers (eg, through a Back button or a Review step which displays a summary of the responses and asks the respondents if they are correct).	Respondents were able to review and change their answers through a Back button.
Response rates			
	Unique site visitor	If you provide view rates or participation rates, you need to define how you determined a unique visitor. There are different techniques available, based on IP addresses or cookies or both.	N/A – Did not provide view rates or participation rates.
	View rate (Ratio of unique survey visitors/unique site visitors)	Requires counting unique visitors to the first page of the survey, divided by the number of unique site visitors (not page views!). It is not unusual to have view rates of less than 0.1 % if the survey is voluntary.	N/A – Survey was advertised through multiple websites, e-mail lists, and social media pages
	Participation rate (Ratio of unique visitors who agreed to participate/unique first survey page	Count the unique number of people who filled in the first survey page (or agreed to participate, for example by checking a checkbox), divided by visitors who visit the first page of the survey (or the informed consents page, if present). This can also be called “recruitment” rate.	Unable to calculate (only know # who visited consent page AND

Item Category	Checklist Item	Explanation	Page #
	visitors)		consented to participate).
	Completion rate (Ratio of users who finished the survey/users who agreed to participate)	The number of people submitting the last questionnaire page, divided by the number of people who agreed to participate (or submitted the first survey page). This is only relevant if there is a separate "informed consent" page or if the survey goes over several pages. This is a measure for attrition. Note that "completion" can involve leaving questionnaire items blank. This is not a measure for how completely questionnaires were filled in. (If you need a measure for this, use the word "completeness rate".)	230 consented to participate and 151 responses were submitted=66% completion rate. Majority of responses not submitted (47/79=59%) were from individuals who consented to participate but never started the survey: Page 7, sentence 1.
<b>Preventing multiple entries from the same individual</b>			
	Cookies used	Indicate whether cookies were used to assign a unique user identifier to each client computer. If so, mention the page on which the cookie was set and read, and how long the cookie was valid. Were duplicate entries avoided by preventing users access to the survey twice; or were duplicate database entries having the same user ID eliminated before analysis? In the latter case, which entries were kept for analysis (eg, the first entry or the most recent)?	N/A – Cookies were not used.
	IP check	Indicate whether the IP address of the client computer was used to identify potential duplicate entries from the same user. If so, mention the period of time for which no two entries from the same IP address were allowed (eg, 24 hours). Were duplicate entries avoided by preventing users with the same IP address access to the survey twice; or were duplicate database entries having the same IP address within a given period of time eliminated before analysis? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	Responses from the same IP address were included regardless of period of time between responses. Those from the same IP address were checked item-by-item to ensure they were not duplicates.
	Log file analysis	Indicate whether other techniques to analyze the log file for identification of multiple entries were used. If so, please describe.	N/A – no log file analysis was performed.
	Registration	In "closed" (non-open) surveys, users need to login first and it is easier to prevent duplicate entries from the same user. Describe how this was done. For example, was the survey never displayed a second time once the user had filled it in, or was the username stored together with the	N/A – open survey.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Item Category	Checklist Item	Explanation	Page #
		survey results and later eliminated? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	
Analysis			
	Handling of incomplete questionnaires	Were only completed questionnaires analyzed? Were questionnaires which terminated early (where, for example, users did not go through all questionnaire pages) also analyzed?	Responses were only analysed if the respondent formally submitted the survey at the end of the last page: Page 6, “Statistical analysis” subheading, sentence 1.
	Questionnaires submitted with an atypical timestamp	Some investigators may measure the time people needed to fill in a questionnaire and exclude questionnaires that were submitted too soon. Specify the timeframe that was used as a cut-off point, and describe how this point was determined.	N/A - timestamps were not evaluated.
	Statistical correction	Indicate whether any methods such as weighting of items or propensity scores have been used to adjust for the non-representative sample; if so, please describe the methods.	N/A – not performed.

STROBE Statement—Checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Page No	Recommendation
Title and abstract	1	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract
		2	(b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction			
Background/rationale	2	3	Explain the scientific background and rationale for the investigation being reported
Objectives	3	4	State specific objectives, including any prespecified hypotheses
Methods			
Study design	4	5	Present key elements of study design early in the paper
Setting	5	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	5	(a) Give the eligibility criteria, and the sources and methods of selection of participants
Variables	7	5; Appendix	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/measurement	8*	5	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	5 - 6	Describe any efforts to address potential sources of bias
Study size	10	N/A	Explain how the study size was arrived at
Quantitative variables	11	6	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	6	(a) Describe all statistical methods, including those used to control for confounding
		N/A	(b) Describe any methods used to examine subgroups and interactions
		6	(c) Explain how missing data were addressed
		N/A	(d) If applicable, describe analytical methods taking account of sampling strategy
		N/A	(e) Describe any sensitivity analyses
Results			
Participants	13*	7	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		7	(b) Give reasons for non-participation at each stage
		N/A	(c) Consider use of a flow diagram
Descriptive data	14*	7	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		7	(b) Indicate number of participants with missing data for each variable of interest
Outcome data	15*	N/A	Report numbers of outcome events or summary measures
Main results	16	7 – 8 Tables 2 - 3	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make

			clear which confounders were adjusted for and why they were included
		N/A	(b) Report category boundaries when continuous variables were categorized
		N/A	(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	N/A	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion			
Key results	18	9	Summarise key results with reference to study objectives
Limitations	19	10	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	9 – 10	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	10	Discuss the generalisability (external validity) of the study results
Other information			
Funding	22	12	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

\*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).